mrp.ase.ro

HUMAN MOBILITY PATTERN MINING: A SYSTEMATIC REVIEW OF METHODS AND DATA PROCESSING

Lizda ISWARI
Mada University, Yogyakarta, Indonesia
and
Universitas Islam Indonesia, Yogyakarta, Indonesia
lizdaiswari@mail.ugm.ac.id

Adhistya Erna PERMANASARI
Gadjah Mada University, Yogyakarta, Indonesia
adhistya@ugm.ac.id

Silmi FAUZIATI
Gadjah Mada University, Yogyakarta, Indonesia
silmi@ugm.ac.id

Abstract
Human mobility pattern mining has emerged as a significant research field, yet existing studies are relatively isolated and lack an integrated review of addressed issues and tested solutions. This systematic review aims to provide a comprehensive analysis of human mobility pattern mining research across three interconnected dimensions: data processing approaches, methodological landscape, and future research directions. Following PRISMA guidelines, this study systematically reviewed 43 carefully selected papers from Scopus-indexed journals, covering the period from 2018 to 2025. A structured two-phase approach was used to extract and categorize information according to four research questions, and then analyze patterns to generate insights and recommendations through comprehensive synthesis. The analysis revealed six distinct conceptual perspectives on human mobility research, five categories of real-world applications, and five common methodological approaches ranging from traditional statistical methods to advanced artificial intelligence techniques. Data quality assessment can be categorized into three fundamental dimensions: completeness, accuracy, and consistency. A four-phase preprocessing pipeline was developed with integrated quality control mechanisms. Current challenges include data quality limitations, temporal dimension inadequacies, and scalability barriers. This review provides a systematic organization of fragmented knowledge and identifies four key future research directions: enhanced data integration, advanced spatiotemporal modeling, semantic enhancement, and scalable computing infrastructure. These findings establish foundations for developing more robust, scalable, and practically applicable mobility analysis frameworks.

Keywords: human mobility, mobility mining, urban mining, trajectory pattern mining, human trajectory

1. INTRODUCTION

Existing research in human mobility pattern mining has developed across multiple directions. However, several critical gaps remain that limit a comprehensive understanding of the field, as previous studies have highlighted the fragmented current research approaches. The existing studies on mobility data are relatively isolated and lack an integrated review of addressed issues and tested solutions (Wang, Miwa and Morikawa, 2020). This isolation has led to inconsistent conceptual foundations, where researchers define human mobility behaviors in different ways without unified theoretical frameworks. Furthermore, the complexity of mobility data processing presents some challenges, as trajectory data without semantic information is hard to understand and interpret (Alowayr *et al.*, 2021), while GPS data often contains outliers and noisy measurements that require cleansing before analysis (Dabbas and Friedrich, 2022).

The methodological approaches reveal additional complexities, with limited methodological knowledge on using geographic context for model generalizability (Roy *et al.*, 2022), and traditional methods lacking accuracy and

ISSN
2067- 2462

Management Research and Practice
Volume 17, Issue 3 / September 2025

granularity in analyzing movements and activities (Huang and Wang, 2022). These methodological limitations are compounded by the lack of standardized evaluation approaches, making it difficult to compare different research methods effectively. The complexity of various networks and pathways for mobility complicates comparisons of research methods (Jahanmanesh, Farhadi and Zamanifar, 2025), while different studies utilize diverse datasets and metrics that further complicate method comparisons.

The need for a comprehensive systematic literature review of human mobility pattern mining research is driven by several critical knowledge gaps that currently limit both theoretical understanding and practical applications. First, data processing approaches lack systematic organization and standardized methodologies. The growth of mobility data sources has created new opportunities but also new challenges in data quality assessment, preprocessing, and integration (Wang, Miwa and Morikawa, 2020). Researchers face difficulties in selecting appropriate data sources and preprocessing techniques for their specific research objectives, as current literature provides limited guidance on these important decisions.

Second, the methodological approaches demonstrate significant diversity without a comprehensive comparative analysis. While this diversity reflects the field's innovation, it also creates confusion about which methods work best for different types of mobility analysis tasks. The absence of systematic validation frameworks makes it difficult for researchers to select appropriate analytical approaches and evaluate their effectiveness.

Finally, rapid technological advances and emerging application domains require systematic identification of innovation trends and future research directions. Current research lacks a comprehensive synthesis of emerging opportunities and persistent challenges that could guide future research development and practical implementation.

This systematic review addresses these gaps by providing a comprehensive analysis across three interconnected dimensions: data processing approaches, methodological approaches, and future research directions. By systematically examining these dimensions, this review contributes to the field by organizing fragmented knowledge, identifying best practices, and providing guidance for both new and experienced researchers in human mobility pattern mining.

The remainder of this paper is organized as follows. Section 2 presents the literature review framework and methodology used for systematic analysis. Section 3 examines the results and discussion organized into two main components: data and preprocessing, covering data source characteristics, quality assessment, preprocessing approaches, and methodological recommendations; and methodological landscape, analyzing common methods and algorithms, validation techniques, and recommended guidelines. Section 4 presents research innovation trends and future directions by discussing methodological innovations, current challenges, and emerging opportunities. Section 5 provides conclusions that synthesize key findings and their implications for advancing human mobility pattern mining research toward more effective, scalable, and practically applicable solutions.

2. SYSTEMATIC REVIEW METHODOLOGY

This study presents a comprehensive literature review of human mobility pattern mining techniques following a previously proposed framework (Wang, Miwa and Morikawa, 2020), as illustrated in Figure 1. The literature review methodology consists of five sequential steps that ensure systematic and comprehensive analysis of the current state of human mobility pattern mining research.

FIGURE 1 - SYSTEMATIC REVIEW FRAMEWORK

Step 1 involves identifying the need for a systematic review, which was accomplished by examining the current state of human mobility pattern mining research and identifying critical knowledge gaps that require comprehensive analysis. The literature reveals several challenges: (1) fragmented conceptual foundations with diverse theoretical perspectives that lack a unified understanding, (2) inconsistent approaches to data processing and quality assessment that limit analytical reliability, (3) varied methodological frameworks without standardized validation approaches, and (4) insufficient synthesis of emerging innovations and future research directions. To address these critical gaps in current understanding, four research questions (RQs) were formulated.

1. **RQ1: What data quality and preprocessing approaches are employed in human mobility pattern mining?** Detailed as (RQ1.1) What are the data quality assessment metrics that are utilized in human mobility pattern mining?; (RQ1.2) What are the preprocessing approaches employed to address data challenges?; and (RQ1.3) What are the recommended preprocessing methodologies for human mobility data?
2. **RQ2: What constitutes the current methodological landscape in human mobility pattern mining?** Detailed as (RQ2.1) What are the common methods and algorithms used for mining human mobility behavior patterns?; (RQ2.2) What validation and evaluation techniques are employed, and how do they align with different mining approaches?; and (RQ2.3): What are the recommended validation and evaluation guidelines for human mobility pattern mining?
3. **RQ3: What are the research innovations, trends, and future directions in human mobility pattern mining?** Detailed as (RQ3.1) What methodological innovations and technological advances are emerging in human mobility pattern mining?; (RQ3.2) What are the key research gaps and persistent challenges that limit current capabilities?; and (R34.3) What future research directions and emerging opportunities can address current limitations?

Step 2 involves identifying a proper review protocol by adopting the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines. A comprehensive search string was constructed combining key terms related to human mobility, analytical methods, and application contexts to capture the breadth of relevant literature:

(("human mobility" OR "mobility pattern" OR "movement pattern*" OR "human movement") AND ("mining" OR "analysis" OR "extraction" OR "clustering" OR "prediction") AND ("urban" OR "city" OR "trajectory" OR "GPS" OR "location"))*

For the inclusion and exclusion criteria, this review focused specifically on outdoor human movement trajectories and pattern mining techniques and excluding non-human movements such as ships, animals, or pollution trajectories. Additional criteria included peer-reviewed journal publications, English language publications, and studies with clear methodological descriptions for human mobility pattern mining.

Step 3 involves searching for primary studies from Scopus-indexed journals covering the period 2018-2025, and the initial search yielded 6,165 publications. After rigorous screening focused on outdoor human movement trajectories and excluding non-human movements, 249 papers were identified for detailed review. The final selection criteria, resulted in 43 highly relevant publications. The temporal distribution of selected papers demonstrates consistent research interest in this field: 2018 (7 papers), 2019 (9 papers), 2020 (3 papers), 2021 (6 papers), 2022 (7 papers), 2023 (4 papers), 2024 (4 papers), and 2025 (4 papers).

Steps 4 and 5 involve extracting relevant information and interpreting findings through a structured two-phase approach as illustrated in Figure 2.

Phase 1: Extract and Categorize Information. We systematically reviewed all 43 papers and extracted key information in relation to each research question. For each research question, information was organized into specific categories that correspond to the sub-questions.

Phase 2: Analyze Patterns and Generate Insights. We analyzed the categorized information to find connections and patterns, then developed implications and recommendations. This process combines findings from different categories to create new knowledge that wasn't obvious from individual papers.

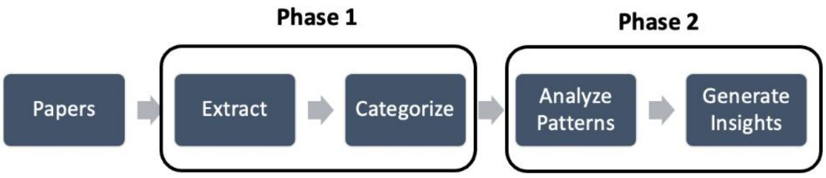

FIGURE 2 - INFORMATION EXTRACTION AND SYNTHESIS METHODOLOGY

As shown in Figure 3, each research question follows a systematic pattern where two foundational sub-questions are combined and analyzed to generate implications and recommendations. For RQ1, data quality assessment and preprocessing approaches are synthesized to create methodology recommendations. For RQ2, methods/algorithms and validation techniques are integrated to develop validation guidelines. The final analytical step involves synthesizing all findings from RQ1 and RQ2 through a comprehensive analysis to address RQ3 about future directions.

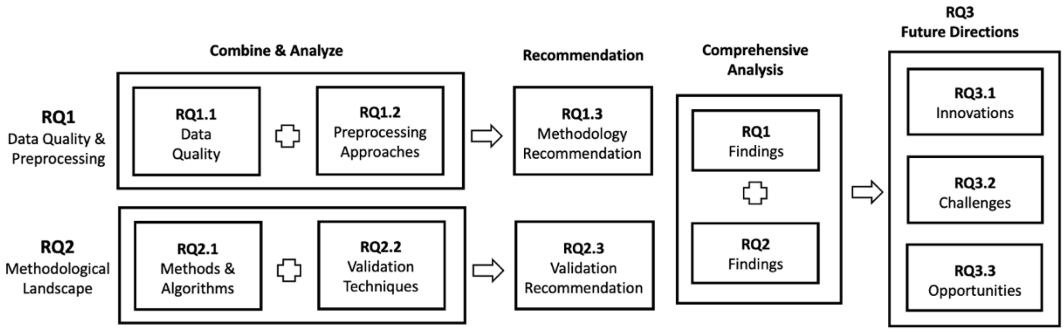

FIGURE 3 - SYSTEMATIC FLOW OF RESEARCH QUESTIONS

3. RESULTS AND DISCUSSION

This section presents comprehensive findings, organized into three sections. The analysis begins with data quality and preprocessing to investigate data quality assessment metrics essential for reliable analysis, preprocessing approaches used to transform raw data into analyzable formats, and methodological recommendations for effective data preparation pipelines. Subsequently, the methodological landscape examines the common algorithms and methods used in mobility pattern mining, as well as validation and evaluation techniques organized by mining approach, and provides comprehensive guidelines for selecting appropriate assessment strategies.

3.1. Data Quality and Preprocessing Approaches

Data quality and preprocessing represent critical components in human mobility pattern mining research, as they determine the quality and reliability of the analytical results. This section examines three essential aspects of data handling in mobility research. First, the types and characteristics of different data sources that

Management Research and Practice
Volume 17, Issue 3 / September 2025

researchers use to capture human mobility behavior. Second, data quality metrics to evaluate data quality before preprocessing and analysis begin. Third, preprocessing approaches that researchers apply to transform raw mobility data into formats that are analyzable. Finally, methodological recommendations for effective preprocessing pipelines based on successful practices across multiple studies.

3.1.1. Types and Characteristics of Data

Human mobility pattern mining research utilizes diverse data sources that capture different aspects of movement behavior. Analysis of the literature reveals six primary data source types with distinct characteristics that define their capabilities and applications across mobility studies.

High-precision data provide exact spatial coordinates and temporal resolution for accurate trajectory reconstruction and location-based investigations. GPS and location data from smartphones, vehicles, and phones are the most popular. Researchers include spatial context with trajectory data from movement monitoring devices, temporal location records, and POI data (Fu *et al.*, 2019; Herberth *et al.*, 2020; Kong *et al.*, 2022; Shan, Sun and Zheng, 2022; Matloub and Kostanic, 2023).

System-level data sources provide transportation network and infrastructure interconnections, revealing system performance and usage patterns. Taxi trip records, fare information, and pickup/dropoff locations reveal urban mobility trends in transportation data. Bus card, metro, and transit schedule data show communal transportation behaviors. License plate recognition (LPR) systems and fleet monitoring data are included in vehicle data, whereas multi-modal transport data helps explain how people navigate urban transportation networks (Zheng *et al.*, 2018; Chen, Cai and Xiong, 2021; Aljeri, 2022; Hussain *et al.*, 2023; Francia, Gallinucci and Golfarelli, 2024; Miao and Liao, 2025).

Population-wide data sources enable research on large demographic groups and geographic areas through data collection infrastructure. Smartphones are used to analyze mobility using Call Detail Records (CDR) and cell tower contacts from telecom companies. Smartphone location data uses cellular location, while sensor data and app usage patterns reveal individual mobility. This data type has population-wide representation and significant period coverage, but poorer spatial precision than GPS data (Thuillier *et al.*, 2018; Fu *et al.*, 2019; Li *et al.*, 2019; Zhu *et al.*, 2019; Seppecher *et al.*, 2021; Solomon *et al.*, 2021; Yin, Lin and Zhao, 2021).

Contextually rich data sources add semantic richness to location data for enhancing data meaning. User-generated location information from geo-tagged tweets from social media platforms provides real-time mobility insights with semantic context. Location metadata in social media posts uniquely combines spatial information with social and semantic elements of mobility behavior to provide rich contextual information about mobility purposes and social aspects of movement (Yang, Cheng and Chen, 2018; Jiao *et al.*, 2019; Zhu *et al.*, 2019; Seppecher *et al.*, 2021; Aljeri, 2022; Shan, Sun and Zheng, 2022).

Real-time data sources enable dynamic applications and responsive systems by detecting and analyzing mobility trends in real time. Sensor data includes Bluetooth proximity detection for co-location and social interaction monitoring. WiFi connection records and access point data provide indoor and urban location information, while mobile device accelerometer and gyroscope data detect transportation modes and activities. IoT sensor networks for traffic monitoring enable large-scale traffic flow analysis and urban mobility comprehension through continuous monitoring and real-time data (Solomon *et al.*, 2021; Jiang *et al.*, 2023; Ibañez *et al.*, 2025; Miao and Liao, 2025).

Ground-truth data sources offer direct verification of mobility behaviors and motivations through user-provided information that cannot be captured through automated methods. Demographic and socioeconomic data explain mobility patterns across population groups, while questionnaire responses provide ground-truth data for validation and a deeper understanding of movement motivations, which validate and explain other data sources (Yin, Lin and Zhao, 2021; Nejadshamsi *et al.*, 2025).

3.1.2. Data Quality Assessment

Effective human mobility pattern mining requires systematic evaluation of data quality before preprocessing and analysis begin. Based on the literature analysis, data quality assessment can be categorized into three

Management Research and Practice
Volume 17, Issue 3 / September 2025

fundamental dimensions that directly impact analytical reliability and research outcomes, as summarized in Table 1.

Completeness Assessment evaluates mobility datasets for gaps or missing values. Time series data continuity is assessed by temporal completeness. Frequent-pattern mining is employed to address gaps in smartphone-based movement data, demonstrating the impact of GPS data gaps on trajectory reconstruction accuracy (Zhao, Jonietz and Raubal, 2021). Sparse datasets require complex reconstruction methods for useful analysis in large-scale low-frequency mobile phone data (Li *et al.*, 2019). Spatial completeness ensures that mobility patterns are captured across the study area without spatial bias. Incomplete trajectories can misinterpret movement patterns and lead to errors in destination identification, as demonstrated by unreliable GPS datasets that present computational efficiency issues (Soares De Sousa, Boukerche and Loureiro, 2023).

Accuracy Assessment compares recorded mobility data with actual movement behaviors for accuracy. Positional accuracy measures the precision of GPS coordinates and the accuracy of measurements. Studies demonstrate that signal interference and multipath effects in urban areas can cause significant positioning errors. Manual validation of 600 waypoints achieved about 98% TMC assignment validation accuracy (Vander Laan, Franz and Marković, 2021). Temporal accuracy assesses the precision and synchronization of timestamps across different data sources, which becomes critical when integrating multiple mobility datasets with varying temporal resolutions. Mobility dataset correlations are investigated using Pearson correlation and t-tests for statistical significance (Chen, Cai and Xiong, 2021).

Consistency Assessment evaluates the uniformity and logical coherence of mobility data across datasets and integrated sources. Internal consistency checks mobility sequences for logical relationships, such as reasonable movement speeds and distances, and chronological order. Chi-square testing is used to verify the consistency of travel hotspots and paths (Du, Meng and Liu, 2024). Integrating diverse data types requires coordinate systems, temporal references, and semantic classifications to be consistent among sources. Format consistency standardizes data structures and measurement units across collecting methods and time periods, addressing sample frequency and data density issues (Chen, Cai and Xiong, 2021).

Practical implementation of quality assessment requires **quantitative metrics** that can guide preprocessing decisions. As described in Table 2, completeness ratios calculate the percentage of complete records versus the total expected records, with thresholds typically set between 70% and 90%, depending on analytical requirements. Accuracy metrics utilize ground truth comparisons where available, with Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) commonly applied for positional accuracy assessment. MAE and standard deviation are used to assess reconstruction quality (Li *et al.*, 2019), while MSE and MAE are applied for validation of behavioral similarity measures (Maiti and Subramanyam, 2019). Consistency validation employs statistical techniques such as outlier detection and logical constraint checking to identify data anomalies that require preprocessing attention.

Preprocessing strategy emerges from quality assessment, enabling researchers to select appropriate preprocessing techniques based on identified data limitations. As summarized in Table 2, high completeness scores may require minimal imputation, whereas low completeness necessitates sophisticated reconstruction methods for trajectory completion (Li *et al.*, 2019) and gap imputation using frequent pattern mining (Zhao, Jonietz and Raubal, 2021). Poor accuracy scores indicate the need for enhanced filtering and smoothing techniques, while consistency issues require standardization and normalization procedures before analysis can proceed effectively. Routing performance assessment and waypoint conflation accuracy provide exemplary models for comprehensive quality evaluation in mobility research (Vander Laan, Franz and Marković, 2021).

TABLE 1 - DATA QUALITY ASSESSMENT

Quality Dimension	Metrics	Description
Completeness Assessment	**Temporal Completeness**	Continuity of time series data, GPS data gaps
	Spatial Completeness	Geographic coverage adequacy across study area
	Trajectory Completeness	Proportion of complete vs. fragmented movement paths
Accuracy Assessment	**Positional Accuracy**	GPS coordinate precision, measurement errors
	Temporal Accuracy	Timestamp precision, synchronization across data sources

	Semantic Accuracy	Correctness of inferred activities and location classifications
Consistency Assessment	**Internal Consistency**	Logical relationships within mobility sequences
	Cross-Source Consistency	Alignment across multiple data types
	Format Consistency	Standardized data structures and measurement units

TABLE 2 - MEASUREMENT APPROACHES

Quantitative Metrics		Preprocessing Strategy	
Completeness Ratios	Percentage of complete records (70-90% thresholds)	**High Completeness**	Minimal imputation needed
		Low Completeness	Sophisticated reconstruction methods required
Accuracy Metrics	MAE, RMSE for positional accuracy	**Poor Accuracy**	Enhanced filtering and smoothing techniques
Consistency Validation	Statistical outlier detection, logical constraint checking	**Consistency Issues**	Standardization and normalization procedures

3.1.3. Data Preprocessing Approaches

Human mobility data preprocessing encompasses a diverse range of techniques that transform raw movement data into analyzable formats. Analysis of the literature reveals several primary preprocessing approaches.

Data Cleaning and Filtering addresses basic data quality problems through systematic removal of problematic records. Key techniques include missing data handling to deal with incomplete datasets (Chen, Cai and Xiong, 2021; Kong *et al.*, 2022; Miao and Liao, 2025), eliminating GPS errors and invalid entries (Chen, Cai and Xiong, 2021; Du, Meng and Liu, 2024), and duplicate elimination to prevent data redundancy (Nejadshamsi *et al.*, 2025). Irrelevant data filtering removes records that do not contribute to analysis objectives (Jiao *et al.*, 2019; Kong *et al.*, 2019).

Semantic Enrichment represents the most common preprocessing approach. This category focuses on adding contextual meaning to raw location data through Points of Interest integration (Yang, Cheng and Chen, 2018; Yin, Lin and Zhao, 2021; Shan, Sun and Zheng, 2022), activity purpose inference (Cai, Lee and Lee, 2018; Yin, Lin and Zhao, 2021), and semantic categorization of locations and movements (Zhu *et al.*, 2019; Shan, Sun and Zheng, 2022).

Clustering and Segmentation group similar data points for analysis. Spatial clustering identifies location-based patterns (Herberth *et al.*, 2020; Solomon *et al.*, 2021; Liu *et al.*, 2022), while temporal segmentation divides data based on time patterns (Zhao *et al.*, 2020; Jiang *et al.*, 2023). Activity clustering groups similar behaviors (Jiao *et al.*, 2019; Yin, Lin and Zhao, 2021), and trajectory clustering organizes movement paths with similar characteristics (Hussain *et al.*, 2023; Si, Yang, Xiang, Wang, *et al.*, 2024).

Temporal Processing handles time-related data features through timestamp adjustment for consistency (Matloub and Kostanic, 2023), time binning and discretization for granularity analysis (Chen, Cai and Xiong, 2021; Aljeri, 2022), and temporal segmentation to identify meaningful time periods (Thuillier *et al.*, 2018; Ma *et al.*, 2024). Time window analysis enables examining mobility patterns within specific time boundaries (Yin, Lin and Zhao, 2021; Jiang *et al.*, 2023).

Trajectory Processing focuses on movement path characteristics through trajectory reconstruction for incomplete paths (Li *et al.*, 2019; Soares De Sousa, Boukerche and Loureiro, 2023), path smoothing to reduce noise (Andrade, Cancela and Gama, 2020), and trip segmentation to find distinct journeys (Zheng *et al.*, 2018; Seppecher *et al.*, 2021). Trajectory compression reduces data size while preserving essential movement information (Yuan *et al.*, 2019).

Spatial Processing handles geographic data characteristics through coordinate system conversion for consistency (Roy *et al.*, 2022), hexagonal binning for spatial analysis (Matloub and Kostanic, 2023), and spatial tessellation for geographic area division (Francia, Gallinucci and Golfarelli, 2024; Hamann and Hagen, 2025).

Stay Point Detection identifies stationary locations through spatio-temporal constraints (Yang, Cheng and Chen, 2018; Zhao, Jonietz and Raubal, 2021), clustering-based detection methods (Solomon *et al.*, 2021; Liu *et al.*, 2022), and duration threshold approaches that define minimum stay times (Matloub and Kostanic, 2023).

Map Matching aligns location data with geographic features through road network alignment (Vander Laan, Franz and Marković, 2021; Hussain *et al.*, 2023), routing engine integration for accurate path reconstruction (Hamann and Hagen, 2025), and network snapping techniques that project GPS points onto transportation networks (Zhao, Jonietz and Raubal, 2021).

3.1.4. Recommendation of Preprocessing Methodology

Based on a comprehensive analysis of preprocessing techniques, general methodological recommendations are proposed to guide future mobility research projects. These recommendations offer practical guidance for implementing effective human mobility preprocessing pipelines. The methodology follows a structured four-phase approach with an integrated activity that ensures systematic and comprehensive data preparation, as illustrated in Figure 6.

Foundation Phase. This phase addresses all basic data quality problems that must be resolved before any sophisticated processing can begin. This comprehensive quality assessment ensures that poor-quality data does not create errors throughout the analytical pipeline. This phase includes d**ata cleaning and filtering**, outlier detection, and noise removal. Data cleaning and filtering handles missing data through appropriate imputation or removal strategies, eliminates erroneous records including GPS errors and invalid coordinates, removes duplicate entries that can skew results, and filters irrelevant data that falls outside the study area.

Standardization Phase. This phase transforms heterogeneous mobility data into consistent formats that enable meaningful analysis across different data sources and collection methods. This includes temporal and spatial processing. **Temporal processing** establishes temporal consistency through timestamp normalization, time zone adjustment, temporal binning, and segmentation. This ensures coherent temporal analysis and prevents inconsistencies due to timing differences rather than actual behavioral variations. **Spatial Processing** involves coordinate system standardization, geographic filtering, spatial discretization, and map matching to align GPS coordinates with transportation networks.

Enhancement Phase. This phase transforms standardized location data into meaningful mobility information by adding semantic context and identifying key behavioral patterns. This includes semantic enrichment and stay point detection. **Semantic Enrichment** adds contextual meaning to raw location data through POI integration, activity inference, semantic categorization, and context enhancement. **Stay Point Detection** identifies stationary locations where individuals spend significant time using spatio-temporal constraints, clustering methods, and duration thresholds. This distinguishes brief stops from meaningful activities and identifies important activity locations.

Refinement Phase. This phase applies final optimization techniques to create clean, complete, and analytically useful trajectory representations. This includes trajectory processing, clustering, and segmentation. **Trajectory processing** optimizes movement path representation through trajectory reconstruction for incomplete paths, path smoothing to reduce irregularities, trip segmentation to identify distinct journeys, and trajectory compression to maintain essential characteristics while reducing data complexity. **Clustering and Segmentation** organize the refined mobility data into meaningful groups through spatial clustering, temporal segmentation, activity clustering, and trajectory clustering to support final pattern analysis.

Validation and Quality Control. This activity encompasses systematic verification of preprocessing results through the data quality assessment framework and continuous monitoring throughout all phases. This applies the three-dimensional assessment approach—completeness, accuracy, and consistency evaluation—at each preprocessing stage to ensure that transformations improve rather than distort underlying mobility patterns. Quantitative metrics guide the selection of preprocessing strategies: completeness ratios (with thresholds of 70-90%) determine whether to use imputation or reconstruction approaches, accuracy measures (MAE, RMSE) trigger the application of filtering and smoothing techniques when needed, and consistency validation through statistical outlier detection identifies the standardization requirements. Combined with sample checking, statistical validation, and comparison with expected mobility patterns, this integrated quality control approach ensures that each preprocessing phase enhances data analytical value while maintaining the integrity of original mobility behaviors.

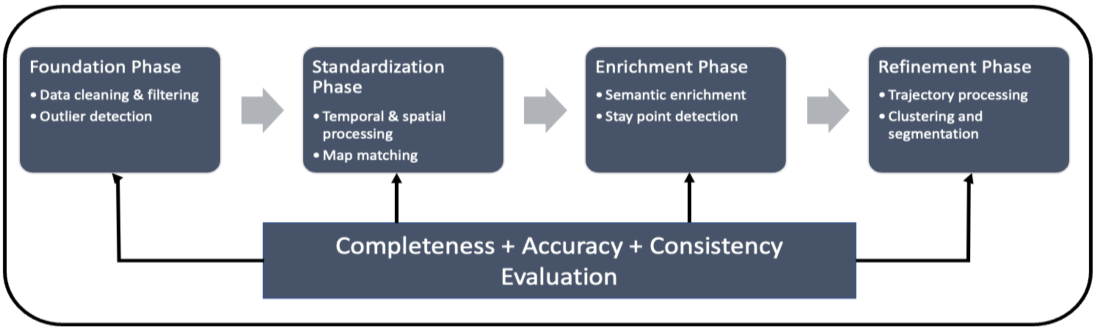

FIGURE 6 - RECOMMENDED PREPROCESSING PIPELINE

3.2. Methodological Landscape

The methodological landscape of human mobility pattern mining represents a diverse and evolving field that combines various analytical approaches. This section examines four key aspects of the methodological framework used in current research. First, we analyze the common methods and algorithms that researchers frequently apply across different mobility studies, categorizing them into five distinct methodological approaches. Second, we explore the validation and evaluation techniques organized by their primary purposes: correctness validation, performance evaluation, and reliability assessment. Third, we examine how validation and evaluation techniques align with specific mining approaches. Finally, we present comprehensive recommendations for selecting appropriate validation and evaluation approaches based on specific analytical tasks in human mobility mining. Together, these components provide a complete methodological framework that guides both method selection and quality assessment in mobility pattern mining research.

3.2.1. Common Methods and Algorithms for Mining Human Mobility

The methodological landscape of mining human mobility encompasses diverse algorithmic approaches that have evolved from traditional statistical methods to sophisticated artificial intelligence techniques. Literature review identifies five common methodological categories, each contributing unique analytical capabilities for extracting meaningful patterns from complex mobility datasets.

Statistical and Spatial Analysis Approaches represent foundational methodologies that utilize mathematical and statistical techniques for analyzing mobility data. These methods include trajectory similarity measures such as Dynamic Time Warping (DTW) and Longest Common Subsequence (LCSS) (Zhu *et al.*, 2019; Jiang *et al.*, 2023; Francia, Gallinucci and Golfarelli, 2024). Spatial parameters, including radius of gyration analysis and cross-correlation techniques (Wijayanto and Wulansari, 2021; Matloub and Kostanic, 2023).

Machine Learning Approaches apply traditional supervised and unsupervised algorithms to discover patterns and relationships within mobility datasets. Classical clustering algorithms, including K-means and DBSCAN, effectively group similar trajectories and identify spatial clusters (Zheng *et al.*, 2018; Aljeri, 2022; Hussain *et al.*, 2023). Advanced techniques, such as Random Forests, Support Vector Machines, and ensemble methods like XGBoost and Gradient Boosting Decision Trees, for classification and prediction tasks (Liu *et al.*, 2022; Roy *et al.*, 2022).

Deep Learning and Neural Network Approaches represent key emerging trends that leverage artificial neural networks to capture complex dependencies and relationships in mobility data. Graph Convolutional Networks (GCNs) effectively handle spatial relationships and network structures inherent in urban transportation systems (Kong *et al.*, 2022). Sequential models, including Long Short-Term Memory (LSTM) networks, process temporal patterns and dependencies in movement sequences (Solomon *et al.*, 2021).

Probabilistic and Stochastic Approaches model mobility patterns as probabilistic processes for uncertainty and variability in human movement behaviors. Hidden Markov Models (HMMs) and Markov chains are relevant for prediction and state transition analysis (Yin, Lin and Zhao, 2021; Soares De Sousa, Boukerche and Loureiro, 2023). These approaches provide robust uncertainty quantification capabilities. They are increasingly integrated with emerging deep learning techniques to create hybrid modeling frameworks that combine probabilistic reasoning with neural network learning.

Pattern Mining and Frequent Pattern Approaches discover recurring patterns and associations within mobility data through analysis of behavioral sequences. Association rule mining, sequential pattern mining, and frequent itemset mining techniques excel at identifying habitual behaviors and routine patterns (Francia, Gallinucci and Golfarelli, 2024; Ma *et al.*, 2024).

3.2.2. Validation and Evaluation Techniques

Validation and evaluation techniques in human mobility pattern mining serve three main purposes that help researchers ensure their methods work correctly and produce reliable results, as shown in Table 3.

Correctness Validation focuses on confirming that methods work as intended and produce accurate results. This includes statistical correlation and significance testing, such as Pearson correlation, t-tests, and chi-square tests, which help researchers determine whether the relationships they discover in mobility data are genuine or merely coincidental. Researchers can validate their findings by conducting ground truth comparisons and cross-validation techniques, including artificial gap testing and survey validation. Pattern discovery and mining validation employ methods such as transportation mode verification and temporal signature validation to ensure that the patterns identified by algorithms accurately represent real mobility behaviors. These correctness validation techniques are essential because they help researchers avoid drawing wrong conclusions from their mobility data analysis.

Performance Evaluation focuses on measuring quantitative performance and accuracy metrics through machine learning metrics, such as MAE, RMSE, F1-score, precision, and recall. These metrics provide researchers with specific numbers to compare the effectiveness of various methods. The Elbow method, Silhouette coefficient, and validity indices are clustering evaluation methods that assist researchers in determining the validity of their clustering results and identifying meaningful groups in mobility data. Accuracy assessment and trajectory reconstruction employ metrics such as reconstruction error and Road Mismatch Fraction to evaluate the precision with which methods can reconstruct movement paths from incomplete data.

Reliability Assessment ensures consistency and robustness across different conditions through algorithm comparison against established baseline methods, cross-validation techniques like k-fold and temporal validation, and robustness testing, including parameter sensitivity and replication studies.

The selection of appropriate validation and evaluation techniques depends on the specific methodological approach employed in human mobility pattern mining, with each approach requiring tailored assessment strategies, as summarized in Table 4.

TABLE 3 - VALIDATION EVALUATION TECHNIQUES BY PURPOSES

Purpose		Type of Validation Evaluation	Example
Correctness Validation	Confirming that methods work as intended and produce accurate results	**Statistical correlation and significance testing**	Pearson correlation, t-tests, chi-square tests
		Ground truth comparison and cross-validation	artificial gap testing, survey validation
		Pattern discovery and mining validation	transportation mode verification, temporal signature validation
Performance Evaluation	Measuring quantitative performance and accuracy metrics	**Machine learning performance metrics**	MAE, RMSE, F1-score, precision, recall
		Clustering evaluation methods	Elbow method, silhouette coefficient, validity indices
		Trajectory reconstruction and accuracy assessment	Road Mismatch Fraction, reconstruction error
Reliability Assessment	Ensuring consistency and robustness across different conditions	**Algorithm comparison and baseline methods**	Comparative analysis against established methods
		Cross-validation techniques	k-fold validation, temporal validation
		Robustness testing	Parameter sensitivity, replication studies

Management Research and Practice
Volume 17, Issue 3 / September 2025

TABLE 4 - VALIDATION/EVALUATION TECHNIQUES BY MINING APPROACH

Validation/Evaluation Technique		Example
Statistical and Spatial Analysis Approaches	Primary Validation: Statistical significance testing (t-tests, chi-square), correlation analysis	Pearson correlation with t-tests (Chen, Cai and Xiong, 2021); Chi-square testing (Du, Meng and Liu, 2024)
	Primary Evaluation: RMSE, MAE for spatial accuracy, Kendall's Tau for ranking accuracy	
Machine Learning Approaches	**Primary Validation**: Cross-validation, ground truth comparison, baseline comparison	F1-score for transportation mode detection (Roy *et al.*, 2022); Precision/recall for activity prediction (Liu *et al.*, 2022)
	Primary Evaluation: F1-score, precision, recall, accuracy metrics, confusion matrices	
Deep Learning and Neural Network Approaches	**Primary Validation**: Train/validation/test split, hyperparameter tuning, ablation studies	MAE and Pearson correlation for flow prediction (Kong *et al.*, 2022); Rand Index for clustering (Si, Yang, Xiang, Li, *et al.*, 2024)
	Primary Evaluation: MAE, Pearson correlation, specialized metrics (UACC, NMI, Rand Index)	
Clustering Approaches	**Primary Validation**: Cluster validity indices, elbow method, visual inspection	Elbow method and validity indices (Aljeri, 2022); Silhouette coefficient (Jiang *et al.*, 2023)
	Primary Evaluation: Silhouette coefficient, compactness measures, separation indices	
Trajectory Analysis Approaches	**Primary Validation**: Manual verification, routing performance validation, map matching accuracy	Manual waypoint validation (Vander Laan, Franz and Marković, 2021); RMF and F1-scores (Soares De Sousa, Boukerche and Loureiro, 2023)
	Primary Evaluation: Road Mismatch Fraction (RMF), trajectory similarity measures, reconstruction error	
Pattern Mining Approaches	**Primary Validation**: Temporal signature verification, transportation mode consistency, survey comparison	Temporal signature validation (Yang, Cheng and Chen, 2018); Transportation mode verification (Zhu *et al.*, 2019)
	Primary Evaluation: Pattern coverage, support measures, semantic consistency	
Probabilistic and Stochastic Approaches	**Primary Validation**: Likelihood testing, model fit assessment, parameter estimation validation	Maximum likelihood validation for distribution fitting
	Primary Evaluation: Log-likelihood, AIC/BIC, prediction accuracy	

3.2.3. Recommended Validation and Evaluation Guidelines

The selection of appropriate validation and evaluation techniques in human mobility pattern mining depends critically on the analytical task and methodological approach employed. As illustrated in Figure 7, there are five common tasks in human mobility mining along with their validation and evaluation recommendation approaches.

For **prediction tasks**, temporal validation strategies that train models on historical data and test on future periods provide the most realistic assessment of predictive capabilities, complemented by cross-validation techniques for robust performance estimation. Performance evaluation relies primarily on regression metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and correlation coefficients that quantify prediction accuracy against observed mobility behaviors.

Classification tasks require stratified cross-validation to maintain class distribution balance across validation folds, with confusion matrix analysis providing detailed insights into classification performance across different mobility categories. The evaluation framework emphasizes F1-score, precision, recall, and ROC-AUC metrics that comprehensively assess classification quality while accounting for potential class imbalances common in mobility data.

Clustering approaches necessitate cluster stability testing and parameter sensitivity analysis to validate the robustness of discovered mobility groups, with evaluation measures like the silhouette coefficient and validity indices, supplemented by visual assessment to ensure meaningful spatial-temporal cluster characteristics.

Management Research and Practice
Volume 17, Issue 3 / September 2025

Trajectory reconstruction and pattern discovery tasks require specialized validation approaches. Trajectory reconstruction validation relies heavily on ground truth comparison using manually verified waypoints or known travel paths, as demonstrated by studies achieving high validation accuracy through systematic manual verification processes. Evaluation metrics focus on spatial accuracy measures, including RMSE for positional accuracy and path similarity measures that assess the fidelity of reconstructed movement sequences.

Pattern discovery tasks employ domain expert validation and survey comparison to ensure that discovered patterns reflect genuine mobility behaviors rather than analytical artifacts, as evidenced by studies comparing mined patterns against national census data and mobility surveys. The evaluation framework emphasizes pattern support measures, coverage metrics that quantify the proportion of mobility data explained by discovered patterns, and semantic consistency measures that assess the meaningfulness of identified patterns within real-world mobility contexts.

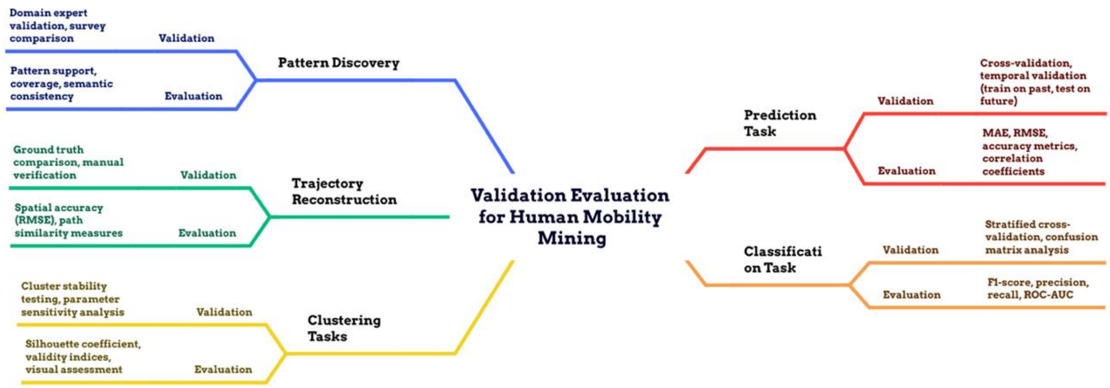

FIGURE 7 - RECOMMENDATION OF VALIDATION AND EVALUATION ON HUMAN MOBILITY TASK

4. RESEARCH INNOVATION, CHALLENGES, AND FUTURE DIRECTIONS IN HUMAN MOBILITY MINING

This section examines three interconnected dimensions: recent methodological innovations, persistent challenges, and future opportunities. Understanding these interconnected aspects provides researchers with a comprehensive roadmap for advancing human mobility pattern mining toward more effective, practical, and beneficial applications.

4.1. Methodological Innovations and Technological Advances

Recent technological innovations in human mobility pattern mining directly address fundamental limitations identified in traditional approaches, representing a paradigm shift toward more sophisticated, accurate, and scalable analytical frameworks. These innovations build upon the foundational statistical and machine learning methods discussed in previous sections while incorporating innovative technologies that overcome limitations in data processing, pattern recognition, and real-time analysis. Methodological innovations can be summarized into four key development areas:

1. **Advanced deep learning and AI technologies** can overcome limitations in complicated pattern recognition and multi-dimensional data analysis that traditional statistical methods could not handle effectively. Multi-Pattern Graph Convolutional Networks (MPGCN) are presented to predict multi-pattern flow using passenger mobility patterns and graph convolutional networks, addressing earlier difficulties in capturing complex urban transportation relationships (Kong *et al.*, 2022). Geographic-Semantic Graph Neural Networks (GSGNN) solve the problem of integrating spatial and semantic information in integrated analytical frameworks by combining geographic close proximity with semantic urban connection (Nejadshamsi *et al.*, 2025).

2. **Multi-modal and context-aware frameworks** improve mobility analysis by combining varied data sources and contextual information, overcoming data integration difficulty and semantic limitations. Semantic itinerary recommenders are developed that combine spatial, temporal, and semantic dimensions from geo-tagged photos, addressing the lack of semantic dimensions in traditional recommender systems (Cai, Lee and Lee, 2018), and geographic context frameworks are created combining GPS features with transportation infrastructure and land use information, overcoming limited geographic context consideration in transportation mode detection (Roy *et al.*, 2022).
3. **Advanced clustering and pattern mining innovations** address computational efficiency and pattern identification constraints in existing methods. GRIDBSCAN and ST-TCLUS are developed for grid density-based and spatiotemporal trajectory clustering, addressing scalability issues in large-scale trajectory analysis (Zheng *et al.*, 2018). In contrast, MIF-STKNNDC is introduced for multidimensional fusion to handle noise, outliers, and complex trajectory distributions (Jiang *et al.*, 2023).
4. **Specialized applications and domain-specific methods** address analytical needs and limits that general-purpose methods cannot handle effectively. Group Movement Pattern Mining based on Similarity (GMPMS) is developed for tourist identification from low-accuracy CDR data to address data sparsity in mobile phone datasets (Zhu *et al.*, 2019). Converging Pattern Mining (CPM) frameworks are developed for proactive incident management to address real-time processing limitations in emergency response applications (Zhao *et al.*, 2020).

4.2. *Current Research Gaps and Persistent Challenges*

Despite significant methodological advances and paradigm shifts described in previous sections, human mobility pattern mining research continues to face persistent challenges that limit both analytical capabilities and practical applications. These challenges represent fundamental barriers that current innovations have not yet fully resolved, requiring continued research attention and novel approaches. The identification of these gaps provides crucial guidance for future research directions and highlights areas where additional innovation is most needed to advance the field toward more robust, scalable, and practically applicable mobility analysis systems.

1. **Data Quality and Scalability Challenges** represent fundamental challenges that affect most mobility research applications. Low sampling rates and data sparsity issues persist in GPS-based studies, resulting in unreliable datasets that pose computational efficiency challenges (Soares De Sousa, Boukerche and Loureiro, 2023). Similarly, low-frequency mobile phone data requires sophisticated reconstruction methods for meaningful analysis (Li *et al.*, 2019). These data quality problems are compounded by GPS data gaps that need specialized approaches such as frequent-pattern mining and time geography constraints to address them (Zhao, Jonietz and Raubal, 2021). Beyond data quality issues, the processing and computational demands of big data create additional scalability barriers. Large-scale social media data shows limited scalability for spatio-temporal analysis (Aljeri, 2022), while network-constrained trajectory clustering requires substantial computational resources (Hussain *et al.*, 2023). These scalability problems highlight the need for distributed computing paradigms for large-scale trajectory mining (Francia, Gallinucci and Golfarelli, 2024).
2. **Methodological and Technical Limitations** reveal gaps in current analytical approaches that prevent a comprehensive understanding of mobility behavior. A major limitation involves temporal dimension inadequacies that persist across many studies. Current clustering methods often neglect temporal dimensions (Si, Yang, Xiang, Li, *et al.*, 2024), research frequently focuses on spatial data while ignoring temporal travel behavior patterns (Maiti and Subramanyam, 2019), and existing network models do not account for temporal passenger travel relationships (Kong *et al.*, 2022). In addition to temporal limitations, semantic and context gaps create further analytical challenges. Raw GPS data lacks semantic information and may contain semantic bias (Shan, Sun and Zheng, 2022), while traditional recommender systems neglect important semantic dimensions that could improve analysis (Cai, Lee and Lee, 2018).

3. **Application-specific and Domain Challenges** create barriers to practical implementation and real-world deployment. Real-time processing represents a significant limitation that prevents dynamic analysis capabilities. Current systems struggle with dynamic traffic conditions and lack real-time adaptability (Miao and Liao, 2025), while researchers encounter difficulties predicting atypical travel behavior and handling varying data density (Herberth *et al.*, 2020). Alongside real-time processing issues, prediction and accuracy problems limit complex pattern recognition capabilities. Multi-pattern passenger flow in complex urban settings cannot be sufficiently captured (Kong *et al.*, 2022), stay point prediction and demographic attribute influence remain challenging (Solomon *et al.*, 2021), and existing methods struggle with noise, outliers, and complex trajectory distributions that are common in real-world mobility data (Jiang *et al.*, 2023).

4.3. *Future Research Directions and Emerging Opportunities*

The analysis of current research gaps and methodological limitations reveals several promising directions for advancing human mobility pattern mining capabilities. Based on a comprehensive analysis of recent developments and identified limitations, future research opportunities can be organized into four interconnected categories.

1. **Enhanced Data Integration and Multi-Source Analytics** represents a critical research direction that addresses current data quality and semantic limitations by integrating diverse data sources. Future research should focus on combining multiple data types, including GPS trajectories, mobile phone records, social media data, IoT sensor networks, and transportation infrastructure data to create comprehensive mobility datasets (Chen, Cai and Xiong, 2021; Aljeri, 2022). Advanced data fusion techniques are needed to handle heterogeneous data formats, varying temporal resolutions, and different spatial accuracies while maintaining analytical consistency across integrated datasets (Fu *et al.*, 2019). Research should also develop standardized preprocessing pipelines that can effectively handle multi-source data integration challenges, including coordinate system alignment, temporal synchronization, and semantic consistency validation (Vander Laan, Franz and Marković, 2021). Additionally, investigation into privacy-preserving data integration methods becomes increasingly important as mobility analysis requires access to sensitive location information from multiple sources (Ibañez *et al.*, 2025).
2. **Advanced Spatiotemporal Modeling and Real-Time Analytics** addresses the temporal dimension inadequacies and real-time processing limitations identified in current research. Future investigations should focus on developing spatiotemporal architectures that can effectively capture both spatial relationships and temporal dynamics in mobility patterns, moving beyond current approaches that primarily emphasize spatial characteristics while neglecting temporal dimensions (Kong *et al.*, 2022; Nejadshamsi *et al.*, 2025). Research opportunities include adaptive graph learning algorithms that can dynamically adjust to changing urban conditions, dynamic attention mechanisms for transportation networks that respond to real-time conditions, and distributed modeling schemes that can handle spatial non-stationarity across different urban contexts (Nejadshamsi *et al.*, 2025). Advanced machine learning architectures, including Transformer models, Generative Adversarial Networks for data augmentation, and hybrid approaches that combine different modeling techniques, should be explored to improve prediction accuracy and handle complex mobility patterns (Ibañez *et al.*, 2025; Miao and Liao, 2025).
3. **Semantic Enhancement and Context-Aware Analytics** represents a fundamental research direction that addresses the semantic and contextual limitations in current mobility analysis approaches. Future research should develop sophisticated methods for semantic trajectory enrichment that can automatically infer activity purposes, transportation modes, and contextual factors from raw movement data (Hamann and Hagen, 2025). Investigation into semi-supervised and unsupervised approaches for semantic annotation becomes critical for handling large-scale datasets where manual labeling is impractical (Shan, Sun and Zheng, 2022). Research should also focus on developing comprehensive validation methodologies for unsupervised learning results in mobility

analysis, as current approaches lack standardized evaluation frameworks (Hamann and Hagen, 2025). Additionally, integration of socioeconomic factors, environmental conditions, and urban infrastructure characteristics into mobility models requires systematic investigation to improve contextual understanding and prediction accuracy (Roy *et al.*, 2022; Nejadshamsi *et al.*, 2025).

4. **Scalable Computing and Technological Infrastructure** addresses the computational and scalability challenges that limit the practical application of current mobility analysis methods. Future research should investigate distributed computing paradigms specifically designed for large-scale trajectory mining, including cloud-based processing architectures and edge computing optimization for real-time mobility analytics (Francia, Gallinucci and Golfarelli, 2024; Miao and Liao, 2025). Development of efficient algorithms that can handle massive datasets while maintaining analytical accuracy represents a critical research priority, particularly for urban areas with high population density and complex transportation networks (Hussain *et al.*, 2023). Research opportunities include exploring blockchain-based security mechanisms for IoT mobility networks, digital twin technology for urban simulation and prediction, and AI-enhanced multi-stage learning approaches for intelligent transportation systems (Miao and Liao, 2025). Investigation into model transferability across different geographic contexts and urban environments is essential for developing universal mobility analysis frameworks that can adapt to diverse conditions without requiring extensive retraining (Roy *et al.*, 2022; Nejadshamsi *et al.*, 2025).

5. CONCLUSIONS

This systematic literature review provides a comprehensive analysis of human mobility pattern mining research across three important dimensions: data processing approaches, methodological landscape, and future research directions. The study reveals that while significant progress has been made in developing sophisticated analytical methods, several fundamental challenges still need to be addressed.

The analysis of 43 carefully selected papers shows that human mobility research has evolved from simple statistical methods to advanced artificial intelligence systems. Based on the literature analysis, data quality assessment can be categorized into three fundamental dimensions that directly impact analytical reliability and research outcomes. However, data quality and scalability issues continue to limit the practical application of many methods. Many studies struggle with incomplete or noisy data, especially when dealing with large-scale datasets. Additionally, most methods have limited ability to work across different geographic contexts or adapt to changing conditions in real-time.

Enhanced data integration and multi-source analytics represent the most needed, requiring a combination of heterogeneous data formats while maintaining analytical consistency. Advanced spatiotemporal modeling and real-time analytics address persistent temporal dimension inadequacies through adaptive graph learning algorithms that can respond to changing urban conditions. Semantic enhancement and context-aware analytics offer solutions for the semantic limitations in current approaches through sophisticated methods for automatic activity purpose inference and contextual factor integration. Scalable computing and technological infrastructure address computational barriers through distributed computing paradigms specifically designed for large-scale trajectory mining and efficient algorithms that maintain analytical accuracy while handling massive datasets.

For new researchers entering this field, this review provides practical guidance on selecting appropriate methods, understanding data requirements, and selecting suitable validation approaches. The systematic organization of knowledge presented here can help researchers avoid common pitfalls and build upon existing work more effectively. Future research should focus on developing methods that are more robust, scalable, and applicable across different contexts while maintaining high standards for privacy protection and the ethical use of mobility data.

REFERENCES

Aljeri, M. (2022) Big Data-Driven Approach to Analyzing Spatio-Temporal Mobility Pattern', *IEEE Access*, 10(September), pp. 98414–98426. Available at: https://doi.org/10.1109/ACCESS.2022.3206859.

Alowayr, A.D. *et al.* (2021) A review of trajectory data mining applications', *2021 International Conference of Women in Data Science at Taif University, WiDSTaif 2021* [Preprint]. Available at: https://doi.org/10.1109/WIDSTAIF52235.2021.9430226.

Andrade, T., Cancela, B. and Gama, J. (2020) From mobility data to habits and common pathways', in *Expert Systems*, pp. 1–20. Available at: https://doi.org/10.1111/exsy.12627.

Cai, G., Lee, K. and Lee, I. (2018) Itinerary recommender system with semantic trajectory pattern mining from geo-tagged photos', *Expert Systems with Applications*, 94, pp. 32–40. Available at: https://doi.org/10.1016/j.eswa.2017.10.049.

Chen, H., Cai, M. and Xiong, C. (2021) Research on human travel correlation for urban transport planning based on multisource data', *Sensors (Switzerland)*, 21(1), pp. 1–16. Available at: https://doi.org/10.3390/s21010195.

Chen, M., Yu, X. and Liu, Y. (2019) TraLFM: Latent Factor Modeling of Traffic Trajectory Data', *IEEE Transactions on Intelligent Transportation Systems*, 20(12), pp. 4624–4634. Available at: https://doi.org/10.1109/TITS.2019.2912075.

Dabbas, H. and Friedrich, B. (2022) Benchmarking machine learning algorithms by inferring transportation modes from unlabeled GPS data', *Transportation Research Procedia*, 62(Ewgt 2021), pp. 383–392. Available at: https://doi.org/10.1016/j.trpro.2022.02.048.

Du, J., Meng, C. and Liu, X. (2024) Analysis of Urban Residents' Travelling Characteristics and Hotspots Based on Taxi Trajectory Data', *Applied Sciences (Switzerland)*, 14(3). Available at: https://doi.org/10.3390/app14031279.

Francia, M., Gallinucci, E. and Golfarelli, M. (2024) Colossal Trajectory Mining: A unifying approach to mine behavioral mobility patterns', *Expert Systems with Applications*, 238(PD), p. 122055. Available at: https://doi.org/10.1016/j.eswa.2023.122055.

Fu, Y. *et al.* (2019) Representing Urban Forms: A Collective Learning Model with Heterogeneous Human Mobility Data', *IEEE Transactions on Knowledge and Data Engineering*, 31(3), pp. 535–548. Available at: https://doi.org/10.1109/TKDE.2018.2837027.

Hamann, J. and Hagen, T. (2025) Revealing Trip Purposes in Raw GPS Data by Applying a Multi-Phase Clustering Approach to Semantic Trajectories', *IEEE Transactions on Intelligent Transportation Systems*, 26(3), pp. 3543–3556. Available at: https://doi.org/10.1109/TITS.2024.3516141.

Herberth, R. *et al.* (2020) Identifying Atypical Travel Patterns for Improved Medium-Term Mobility Prediction', *IEEE Transactions on Intelligent Transportation Systems*, 21(12), pp. 5010–5021. Available at: https://doi.org/10.1109/TITS.2019.2947347.

Huang, W. and Wang, L. (2022) Towards big data behavioral analysis: rethinking GPS trajectory mining approaches from geographic, semantic, and quantitative perspectives', *Architectural Intelligence*, 1(1), pp. 1–15. Available at: https://doi.org/10.1007/s44223-022-00011-y.

Hussain, S.A. *et al.* (2023) Efficient Trajectory Clustering with Road Network Constraints Based on Spatiotemporal Buffering', *ISPRS International Journal of Geo-Information*, 12(3). Available at: https://doi.org/10.3390/ijgi12030117.

Ibañez, G.C. *et al.* (2025) MobilitApp: A Deep Learning-Based Tool for Transport Mode Detection to Support Sustainable Urban Mobility', *IEEE Access*, 13(March), pp. 68439–68461. Available at: https://doi.org/10.1109/ACCESS.2025.3561238.

Jahanmanesh, P., Farhadi, A. and Zamanifar, A. (2025) Mobility Management With AI', *IEEE Access*, 13(May), pp. 83022–83060. Available at: https://doi.org/10.1109/ACCESS.2025.3566337.

Jiang, Q. *et al.* (2023) Behavior pattern mining based on spatiotemporal trajectory multidimensional information fusion', *Chinese Journal of Aeronautics*, 36(4), pp. 387–399. Available at: https://doi.org/10.1016/j.cja.2022.10.010.

Jiao, X. *et al.* (2019) Exploring Spatial and Mobility Pattern's Effects for Collaborative Point-of-Interest Recommendation', *IEEE Access*, 7, pp. 158917–158930. Available at: https://doi.org/10.1109/ACCESS.2019.2950927.

Kong, X. *et al.* (2019) COOC: Visual Exploration of Co-Occurrence Mobility Patterns in Urban Scenarios', *IEEE Transactions on Computational Social Systems*, 6(3), pp. 403–413. Available at: https://doi.org/10.1109/TCSS.2018.2883582.

Kong, X. *et al.* (2022) Exploring Human Mobility for Multi-Pattern Passenger Prediction: A Graph Learning Framework', *IEEE Transactions on Intelligent Transportation Systems*, 23(9), pp. 16148–16160. Available at: https://doi.org/10.1109/TITS.2022.3148116.

Vander Laan, Z., Franz, M. and Marković, N. (2021) Scalable Framework for Enhancing Raw GPS Trajectory Data: Application to Trip Analytics for Transportation Planning', *Journal of Big Data Analytics in Transportation*, 3(2), pp. 119–139. Available at: https://doi.org/10.1007/s42421-021-00040-5.

Li, M. *et al.* (2019) Reconstruction of human movement trajectories from large-scale low-frequency mobile phone data', *Computers, Environment and Urban Systems*, 77(May), p. 101346. Available at: https://doi.org/10.1016/j.compenvurbsys.2019.101346.

Liu, X. *et al.* (2022) Graph-based representation for identifying individual travel activities with spatiotemporal trajectories and POI data', *Scientific Reports*, 12(1), pp. 1–13. Available at: https://doi.org/10.1038/s41598-022-19441-9.

Ma, C. *et al.* (2024) Understanding Residents ' Behavior for Smart City Management by Sequential and Periodic', *IEEE Transactions on Computational Social Systems*, 11(1), pp. 1260–1276. Available at: https://doi.org/10.1109/TCSS.2023.3249740.

Maiti, S. and Subramanyam, R.B.V. (2019) Mining behavioural patterns from spatial data', *Engineering Science and Technology, an International Journal*, 22(2), pp. 618–628. Available at: https://doi.org/10.1016/j.jestch.2018.10.007.

Matloub, Z. and Kostanic, I. (2023) Comprehensive Analysis of Key Spatial Parameters for Characterizing Human Mobility Patterns', *IEEE Access*, 11(September), pp. 106232–106248. Available at: https://doi.org/10.1109/ACCESS.2023.3320572.

Miao, Z. and Liao, Q. (2025) IoT-Based Traffic Prediction for Smart Cities', *IEEE Access*, 13(March), pp. 52369–52384. Available at: https://doi.org/10.1109/ACCESS.2025.3552276.

Nejadshamsi, S. *et al.* (2025) A geographic-semantic context-aware urban commuting flow prediction model using graph neural network', *Expert Systems with Applications*, 261(January 2024), p. 125534. Available at: https://doi.org/10.1016/j.eswa.2024.125534.

Roy, A. *et al.* (2022) Assessing the role of geographic context in transportation mode detection from GPS data', *Journal of Transport Geography*, 100(March), p. 103330. Available at: https://doi.org/10.1016/j.jtrangeo.2022.103330.

Seppecher, M. *et al.* (2021) Estimation of urban zonal speed dynamics from user-activity-dependent positioning data and regional paths', *Transportation Research Part C: Emerging Technologies*, 129(October 2020), p. 103183. Available at: https://doi.org/10.1016/j.trc.2021.103183.

Shan, Z., Sun, W. and Zheng, B. (2022) Extract Human Mobility Patterns Powered by City Semantic Diagram', *IEEE Transactions on Knowledge and Data Engineering*, 34(8), pp. 3765–3778. Available at: https://doi.org/10.1109/TKDE.2020.3026235.

Si, J., Yang, J., Xiang, Y., Li, L., *et al.* (2024) ConDTC: Contrastive Deep Trajectory Clustering for Fine-grained Mobility Pattern Mining', *IEEE Transactions on Big Data*, 14(8). Available at: https://doi.org/10.1109/TBDATA.2024.3362195.

Si, J., Yang, J., Xiang, Y., Wang, H., *et al.* (2024) TrajBERT: BERT-Based Trajectory Recovery With Spatial-Temporal Refinement for Implicit Sparse Trajectories', *IEEE Transactions on Mobile Computing*, 23(5), pp. 4849–4860. Available at: https://doi.org/10.1109/TMC.2023.3297115.

Soares De Sousa, R., Boukerche, A. and Loureiro, A.A.F. (2023) A Novel Scalable Framework to Reconstruct Vehicular Trajectories from Unreliable GPS Datasets', *IEEE Transactions on Intelligent Transportation Systems*, 24(9), pp. 9658–9669. Available at: https://doi.org/10.1109/TITS.2023.3280114.

Solomon, A. *et al.* (2021) Analyzing movement predictability using human attributes and behavioral patterns', *Computers, Environment and Urban Systems*, 87(January 2020). Available at: https://doi.org/10.1016/j.compenvurbsys.2021.101596.

Thuillier, E. *et al.* (2018) Clustering Weekly Patterns of Human Mobility Through Mobile Phone Data', *IEEE Transactions on Mobile Computing*, 17(4), pp. 817–830. Available at: https://doi.org/10.1109/TMC.2017.2742953.

Wang, D., Miwa, T. and Morikawa, T. (2020) *Big trajectory data mining: A survey of methods, applications, and services*, *Sensors (Switzerland)*. Available at: https://doi.org/10.3390/s20164571.

Wang, P. *et al.* (2018) Learning urban community structures: A collective embedding perspective with periodic spatial-temporal mobility graphs', *ACM Transactions on Intelligent Systems and Technology*, 9(6). Available at: https://doi.org/10.1145/3209686.

Wijayanto, A.W. and Wulansari, I.Y. (2021) Human Mobility Patterns and Its Cross-Correlation with the COVID-19 Transmission in Jakarta, Indonesia', *Journal of Physics: Conference Series*, 1863(1). Available at: https://doi.org/10.1088/1742-6596/1863/1/012017.

Yang, M., Cheng, C. and Chen, B. (2018) Mining individual similarity by assessing interactions with personally significant places from GPS trajectories', *ISPRS International Journal of Geo-Information*, 7(3). Available at: https://doi.org/10.3390/ijgi7030126.

Yin, L., Lin, N. and Zhao, Z. (2021) Mining Daily Activity Chains from Large-Scale Mobile Phone Location Data', *Cities*, 109(October 2020). Available at: https://doi.org/10.1016/j.cities.2020.103013.

Yuan, G. *et al.* (2019) APDS: A framework for discovering movement pattern from trajectory database', *International Journal of Distributed Sensor Networks*, 15(11). Available at: https://doi.org/10.1177/1550147719888164.

Zhao, B. *et al.* (2020) A Framework for Group Converging Pattern Mining using Spatiotemporal Trajectories', *GeoInformatica*, 24(4), pp. 745–776. Available at: https://doi.org/10.1007/s10707-020-00404-z.

Zhao, P., Jonietz, D. and Raubal, M. (2021) Applying frequent-pattern mining and time geography to impute gaps in smartphone-based human-movement data', *International Journal of Geographical Information Science*, 35(11), pp. 2187–2215. Available at: https://doi.org/10.1080/13658816.2020.1862126.

Zheng, L. *et al.* (2018) Spatial–temporal travel pattern mining using massive taxi trajectory data', *Physica A: Statistical Mechanics and its Applications*, 501, pp. 24–41. Available at: https://doi.org/10.1016/j.physa.2018.02.064.

Zhu, X. *et al.* (2019) Exploring group movement pattern through cellular data: A case study of tourists in Hainan', *ISPRS International Journal of Geo-Information*, 8(2). Available at: https://doi.org/10.3390/ijgi8020074.